ESTADÍSTICA ESPAÑOLA
Vol. 41, Núm. 144, 1999, págs. 47 a 78

Una comparación entre métodos de Estimación de Factores Comunes y Relaciones de Cointegración en Vectores de Series Temporales (1)

por
LUIS MANZANEDO
Oficina Nacional de Inspección de la Agencia Tributaria
Ministerio de Economía y Hacienda

DANIEL PEÑA
Departamento de Estadística y Econometría
Universidad Carlos III de Madrid

RESUMEN

Este trabajo presenta una comparación empírica de los métodos de Box y Tiao (1977), Peña y Box (1987) y Johansen (1988) para la estimación de factores comunes y cointegrados en vectores de series temporales. Los tres métodos proporcionan buenos resultados para la estimación de factores comunes I(1), sin embargo, el error de estimación aumenta en el caso de I(2), y en conjunto el mejor procedimiento es el de Peña y Box (1987). En la estimación del coeficiente de cointegración el mejor método es el de Johansen, si bien, cuando el proceso generador de datos tiene deriva los tres métodos presentan propiedades similares.

(1) Esta versión se ha beneficiado de los comentarios de dos evaluadores anónimos, a los cuales queremos expresar nuestra gratitud.

Palabras clave: Vectores propios, transformaciones lineales, series temporales no estacionarias.

Clasificación AMS: 90A20, 62M20, 62P20.

1. INTRODUCCIÓN

Engle y Granger (1987) mostraron que aunque todos los componentes de un vector de series temporales sean no estacionarios pueden existir combinaciones lineales estacionarias de estos componentes que se denominan relaciones de cointegración. Su estudio y estimación han sido objeto de numerosos trabajos, incluyendo Johansen (1988, 1995), Stock y Watson (1988, 1993) y Chan y Tsay (1991). Hargreaves (1994) presenta una buena panorámica de este problema que está estrechamente relacionado con la presencia de factores comunes en vectores de series temporales. Quenouille (1957) fue pionero en el tratamiento de raíces unitarias en vectores de series temporales con estructura autorregresiva vectorial de primer orden (VAR(1)). Posteriormente, Box y Tiao (1977), Peña y Box (1987), Tiao y Tsay (1989), Reinsel y Ahn (1992), Escribano y Peña (1994) y Gonzalo y Granger (1995) han realizado aportaciones a este campo.

En los últimos años se han realizado diversos trabajos de simulación para comparar la eficiencia de distintos métodos para estimar las relaciones de cointegración. Entre ellos hay que destacar los realizados por Gonzalo (1994) y Bewley y otros (1994). Este trabajo contribuye a esta literatura comparando mediante un experimento de Monte Carlo los métodos para determinar factores comunes de Box y Tiao (1977), de Peña y Box (1987), y el método de obtención de las relaciones de cointegración debido a Johansen (1988).

El resto del artículo está estructurado como sigue. En la sección 2 revisamos brevemente la relación entre los conceptos de factores comunes y cointegración. En la sección 3 presentamos los métodos de estimación que van a ser objeto de comparación en este trabajo. En la sección 4 describimos los resultados de un estudio de Monte Carlo para comparar los factores comunes obtenidos por los tres métodos de estimación de Box y Tiao, Peña y Box y Johansen. En la sección 5 comparamos la estimación por los tres métodos del coeficiente de cointegración en una relación. Finalmente, en la sección 6 presentamos las conclusiones del trabajo.

2. EL MODELO FACTORIAL Y COINTEGRACIÓN

Sea X^*_t un vector de n series temporales que, para simplificar la exposición, transformamos a desviaciones respecto a cierto origen μ_t , $X_t = X^*_t - \mu_t$. Suponemos como condiciones iniciales $X^*_j = 0$ para $j \leq 0$. Si el vector X^*_t es estacionario tomamos $\mu_t = \mu = E(X^*_t)$. En otro caso, μ_t es un vector columna de n componentes que puede incluir constantes, tendencias deterministas, variables ficticias etc.

La experiencia empírica indica que es frecuente encontrar relaciones entre los componentes de un vector de series temporales que presentan un comportamiento más estable que los componentes individuales. Hay dos enfoques principales para modelar esta situación: suponer factores comunes o suponer relaciones de cointegración. Peña y Box (1987) proponen un modelo factorial dinámico para un vector de n series temporales X_t estacionarias mediante las relaciones :

$$X_t = Af_t + u_t, \qquad [1]$$

$$\Phi(B)f_t = \Theta(B)a_t, \qquad [2]$$

donde A es una matriz (nxr) de rango completo, la matriz de carga, que contiene los coeficientes de la relación entre las series observadas y los r factores comunes, f_t , que siguen un modelo VARMA, y a_t y u_t son procesos de ruido blanco independientes, con media nula, matrices de varianzas-covarianzas Σ_a y Σ_u, respectivamente de rango completo. Se supondrá que todas las matrices de la ecuación de generación de los factores, Φ y Θ y la matriz Σ_a son diagonales y, para asegurar la identificabilidad del modelo, impondremos que $A'A=I$. Estas condiciones pueden relajarse para permitir factores dependientes, pero para simplificar la exposición en este trabajo consideramos el caso de factores no correlacionados. Este modelo ha sido generalizado por Peña y Poncela (1998) al caso no estacionario. Un modelo alternativo al factorial es la formulación autorregresiva de rango reducido, (véase Reinsel, 1993) donde se supone que las matrices de la formulación VAR de un vector de series temporales tienen rango menor que n. Este modelo puede considerarse un caso particular del modelo factorial, por lo que no entraremos en su desarrollo.

El segundo enfoque para explicar la estabilidad de ciertas relaciones lineales entre los componentes de una serie es el concepto de cointegración. Diremos que una serie temporal univariante X_t es integrada de orden d, I(d), si $(1-B)^d X_t$ sigue un proceso MA(∞) invertible . Por otro lado, el vector de series X_t es conjuntamente integrado de orden d, JI(d) si: (1) todos los componentes son I(d); (2) $(1-B)^d X_t$ sigue un vector VMA(∞) invertible de rango completo. Finalmente, los componentes del vector X_t están cointegrados de orden b, CI(d,b) con rango s si: (1) todos los componentes son I(d); (2) existen s combinaciones lineales independientes entre sí, $\beta' X_t$, que son JI(d-b). La matriz (nxs) β se denomina matriz de cointegración.

Escribano y Peña (1994) demuestran que el enfoque de factores comunes y el de cointegración son formas alternativas para el análisis de vectores de series temporales. En concreto, para el caso particular de procesos integrados I(1), prueban la equivalencia de las tres proposiciones siguientes:

a.- X_t es integrada de orden 1, I(1), con los componentes cointegrados de orden 1, CI(1,1), con rango s;

b.- X_t puede escribirse como generada por r = n-s tendencias comunes;

c.- Xt tiene r factores comunes que son conjuntamente integrados de orden 1, JI(1), y s factores cointegrados que son JI(0).

La relación entre el modelo factorial y el de cointegración es intuitivamente esperable a partir de la formulación [1] del modelo factorial. Supongamos factores no estacionarios que crean la no estacionaridad de la serie y que siguen un paseo aleatorio multivariante, es decir, los factores son I(1), y que los componentes específicos son ruido blanco como indica la ecuación [1]. Sea β la matriz nx(n-r) complemento ortogonal de A, es decir β' A=0. Entonces multiplicando la ecuación [1] por β' tenemos que :

$$\beta' X_t = \beta' A f_t + \beta' u_t = \beta' u_t,$$

y obtenemos combinaciones lineales estacionarias. Esta expresión indica que la matriz de cointegración, β, puede obtenerse como el complemento ortogonal de la de carga y viceversa.

3. MÉTODOS DE ESTIMACIÓN

En esta sección vamos a estudiar tres métodos para obtener la matriz de carga del modelo factorial y su complemento ortogonal, la matriz de cointegración: el análisis canónico de Box y Tiao (1977), el modelo de Factores Comunes de Peña y Box (1987) y Peña y Poncela (1998), y el método de Máxima Verosimilitud de Johansen (1988).

3.1. El análisis canónico de Box y Tiao

Box y Tiao (1977) efectúan un análisis de la estructura subyacente de un vector de series X_t estacionario, de media nula y con representación VAR(p), sobre la base del poder predictivo de su pasado. Partiremos de la descomposición:

$$X_t = F_{t-1}(1) + \eta_t, \qquad [3]$$

donde $F_{t-1}(1)=E(X_t / X_{t-1}..., X_1)$ es la predicción de mínimo error cuadrático medio de X_t en el período t-1 (la proyección ortogonal de X_t sobre el espacio generado por sus valores pasados) y η_t el error asociado a esta predicción. Sea $\Gamma_X = E[X_t X'_t]$ la matriz de covarianzas del vector observado, $C_X = E[F_{t-1}(1)\ F_{t-1}(1)']$ la matriz de covarianzas de la parte predecible o explicada y Σ la matriz de covarianzas del error de predicción a un paso, o variabilidad no explicada. Suponemos que las matrices Γ_X y Σ son definidas positivas.

La ecuación [3] permite escribir una descomposición de la varibilidad de forma análoga al análisis de la varianza multivariante:

$$\Gamma_X = C_X + \Sigma,$$

y si consideramos una combinación lineal $y_t = m'X_t$ es fácil comprobar que el vector m que maximiza el cociente entre la variabilidad explicada y la no explicada es el mayor vector propio de la matriz $\Lambda_X = \Gamma_X^{-1} C_X$. Por tanto, sean $\lambda_1 > \lambda_2 > ..... > \lambda_n$ los valores propios ordenados de esta matriz y sean $m_1, m_2, ..., m_n$ los correspondientes autovectores. Estos autores demuestran que si formamos la matriz M que tiene por columnas los autovectores $m_1, m_2, ..., m_n$, las matrices $M'\Gamma_X M$, $M'C_X M$ y $M'\Sigma M$ son las tres diagonales. Entonces, si definimos el nuevo vector temporal $Y_t = M'X_t$, sus componentes (1) quedan ordenados de mayor a menor predictibilidad; (2) son contemporáneamente independientes; (3) generan predicciones que son independientes en cada instante.

Si los componentes de X_t no son estacionarios las matrices de covarianzas no están definidas, y el método anterior no podría aplicarse. Box y Tiao demuestran que si el proceso tiene r raices en la parte AR que tienden a la unidad, r valores propios de Λ_X serán iguales a la unidad. Para generalizar este resultado podemos utilizar en lugar de las matrices de autocovarianzas las matrices de covarianzas generalizadas (Peña y Poncela, 1998). Estas matrices se definen dividiendo las sumas de cuadrados por una constante adecuada que asegure la convergencia de la matriz. Por ejemplo, para un proceso I(1) la matriz de covarianzas generalizada es Γ_X /N. De esta manera estas matrices convergen a matrices aleatorias cuyos términos son funcionales de procesos de Wiener, (Chang y Wei, 1988). En particular, la matriz $\Lambda_X = \Gamma_X^{-1} C_X$ convergerá a una matriz aleatoria con propiedades similares al caso estacionario. Si existen n-r factores cointegrados entre las variables, obtendremos n-r combinaciones lineales estacionarias por los vectores propios correspondientes a los valores propios pequeños de esta matriz y r combinaciones no estacionarias correspondientes a los vectores propios de valor aproximadamente la unidad.

Aunque el método anterior no está en principio diseñado para estimar la matriz de carga de un modelo factorial puede adaptarse fácilmente para este cometido. Observemos que como la matriz Λ_X no es simétrica, sus vectores propios no son necesariamente ortogonales, aunque puede demostrarse fácilmente que sus valores propios son reales. Supongamos que hemos obtenido s=n-r factores JI(0), y sea β la matriz que contiene los vectores propios ligados a valores propios distintos de uno que han producido estas series estacionarias. Claramente esta matriz es una estimación de la matriz de cointegración. Podemos encontrar su complemento ortogonal y asociarlo a los factores comunes. Para ello observemos que las matrices $C_X \Gamma_X^{-1}$ y $\Gamma_X^{-1}C_X$ tienen los mismos autovalores y que los autovectores M de la matriz $\Gamma_X^{-1}C_X$ están relacionados con los autovectores V de la matriz $C_X \Gamma_X^{-1}$ por $M = \Gamma_X^{-1}V$. Este resultado es la base del siguiente:

Teorema. El complemento ortogonal de la matriz de cointegración $\beta = (m_{r+1},....,m_n)$, formada por los autovectores asociados a los autovalores distintos de la unidad de la matriz $\Gamma_X^{-1}C_X$ que conducen a series estacionarias, es la matriz $\beta_\perp = (v_1,....,v_r)$ que contiene los autovectores de la matriz $C_X \Gamma_X^{-1}$ ligados al autovalor unidad.

Demostración

Sea la matriz de cointegración $\beta = (m_{r+1},....,m_n)$. Debemos encontrar una matriz $\beta_\perp$ tal que $(\beta_\perp\ \beta)$ sea de rango completo y $\beta_\perp'\beta = 0$. La condición $M'\Gamma_X M$ = Diagonal implica que $(m_1,....,m_r)'\Gamma_X\beta = 0$, tomemos $\beta_\perp = \Gamma_X(m_1,....,m_r)$, y por la relación entre los autovectores de las matrices $C_X \Gamma_X^{-1}$ y $\Gamma_X^{-1}C_X$, se tendra que $\beta_\perp = (v_1,....,v_r)$; siendo $v_1,....,v_r$ los autovectores asociados a los r mayores autovalores de la matriz $C_X\Gamma_X^{-1}$.

Concluimos, pues, que la matriz de factores comunes del vector temporal X_t, asociada al método de Box y Tiao (1977), puede obtenerse mediante los autovectores asociados a los r autovalores mayores (próximos a uno) de la matriz $C_X \Gamma_X^{-1}$.

3.2. Factores Comunes de Peña y Box

Peña y Box (1987) demuestran que para series estacionarias que siguen el modelo factorial [1] las matrices de covarianzas tienen la estructura, para k>0,

$$\Gamma_X(k) = AD(k)A',$$

donde A es la matriz de carga y D(k) es una matriz diagonal definida positiva. Las columnas de A son vectores propios de las matrices de covarianzas retardadas con valores propios los términos diagonales de D(k). Partiendo de este resultado, estos autores proponen obtener una estimación preliminar de la matriz de carga, en un vector de series temporales estacionarias, a partir de los vectores propios comunes de las matrices de covarianzas retardadas. Obtenida la matriz A podemos obtener fácilmente la de cointegración como el complemento ortogonal de A. Para ello formamos la matriz AA' y nos quedamos con los vectores propios ligados a valores propios nulos. Estos vectores estiman la matriz de cointegracion β. Para obtener la transformación que separa los factores de un vector de ruido blanco definimos la matriz $M=[A\ \beta]$, y la transformación $Y_t = M'X_t$ nos proporcionará en las primeras r componentes los factores mezclados con ruido y en las n-r siguientes componentes ruido blanco.

Cuando el vector analizado no es estacionario la matriz de covarianzas no existe y, en consecuencia, no podrá aplicarse el método de Peña y Box (1987) para la extracción de los factores comunes. Peña y Poncela (1998) solucionan el problema analizando las matrices de covarianzas generalizadas. Supongamos para simplificar que el vector de series X_t es I(1). Entonces, utilizando [1] podemos escribir:

$$N^{-2}\sum X_t X'_{t-k} = N^{-2}A\sum f_t f'_{t-k}A' + N^{-2}\sum u_t u'_{t-k} + N^{-2}\sum(u_t f'_{t-k}A' + Af_t u'_{t-k}). \quad [4]$$

Las matrices $N^{-2}\sum X_t X'_t$, $N^{-2}\sum f_t f'_t$ convergen a sendas matrices aleatorias cuyos términos son funcionales de procesos de Wiener, mientras que el resto de los términos converge hacia cero. En consecuencia, para procesos I(1) existe una descomposición análoga al caso estacionario y obteniendo los valores y vectores propios de las matrices de covarianzas generalizadas $C(k)=N^{-2}\sum X_t X'_{t-k}$, podremos efectuar la descomposición entre factores estacionarios y no estacionarios. Basta observar que valores propios son constantes en las matrices y tomar éstos para definir la matriz de carga. Dado que los valores propios variarán ligeramente de una matriz a otra en la práctica tomaremos los valores propios de C(1).

Llamando A a la matriz de los autovectores asociados a los r autovalores mayores y repetidos en las matrices C(k) tal y como aparecen en la matriz C(1) y β a la matriz de los n-r autovectores restantes de esta matriz, tendremos que:

a.- existirán n-r factores cointegrados, dados por $c_t = \beta'X_t$;

b.- existirán r factores comunes que pueden estimarse por $f_t = A'X_t$;

c.- los factores comunes serán ortogonales a los cointegrados.

Por tanto, con esta metodología se obtiene una estimación inicial de la matriz de carga y de cointegración, así como una transformación para recuperar los factores. Una ventaja adicional de este planteamiento es que, a diferencia del análisis canónico, el modelo factorial supone un modelo formal que permite la estimación máximo verosímil de la matriz de carga escribiendo el modelo en el espacio de los estados y utilizando el algoritmo EM (Peña y Poncela, 1998). Sin embargo, en el resto de la exposición supondremos que los factores se estiman por el método anterior.

3.3. Método de Máxima Verosimilitud de Johansen

Johansen (1988) propone la estimación por máxima verosimilitud de la matriz de cointegración asociada a un vector de series temporales I(1), X_t, en el que existen r factores comunes. Suponiendo que el vector X_t sigue un proceso VAR(p), el procedimiento de Johansen se reduce a realizar un análisis de correlación canónica entre los residuos de las regresiones de $(1\text{-}B)X_t$ y X_{t-p} sobre $(1\text{-}B)X_{t-1}.....(1\text{-}B)X_{t-p+1}$. No entraremos en la descripción del método porque el lector interesado puede encontrarlo en cualquier texto de econometría moderno (véase por ejemplo Banarjee y otros, 1993). La matriz de carga para los factores comunes se estimará como el complemento ortogonal de la matriz de cointegración estimada por Johansen.

4. COMPARACIÓN MEDIANTE SIMULACIONES DE LOS MÉTODOS DE ESTIMACIÓN DE LA MATRIZ FACTORIAL

En esta sección vamos a comparar mediante un experimento de simulación los tres métodos estudiados de estimación de factores comunes. El experimento consistirá en generar un vector de series mediante un modelo factorial con matriz de carga A y estimar esta matriz con los tres métodos anteriores. Un problema a la hora de definir un criterio de comparación es que cuando definimos el subespacio que contiene las columnas de la matriz de carga, que llamaremos espacio de los factores, cualquier base dentro de este espacio es una solución. Como el modelo factorial está indeterminado ante rotaciones, cualquier otra base de ese espacio es una solución del problema. Esto implica que no podemos medir la discrepancia mediante una función de la diferencia entre la matriz factorial A y su estimación A^* ya que esta diferencia no es invariante ante rotaciones. Si aplicamos una transformación ortogonal R, tal que $H = AR$ con $RR' = I$, se verifica que $H–H^* \neq A–A^*$. Sin embargo, observemos que la matriz:

$$HH' - H^* H^{*\prime} = ARR'A' - A^* RR' A^{*\prime} = AA' - A^* A^{*\prime},$$

si es invariante ante rotaciones. En consecuencia, una medida de discrepancia entre ambos subespacios con el criterio del error cuadrático medio puede definirse mediante:

$$ECM=1' (AA' - A^* A^{*\prime})1, \quad [5]$$

donde 1 representa un vector de unos. Vemos que este criterio tiene en cuenta las diferencias entre los módulos y los ángulos de los vectores fila que forman la matriz de carga. Otro criterio posible es medir el ángulo entre ambos subespacios, lo que equivale a obtener el mayor valor propio de la matriz $A' A^* A^{*\prime} A$. Como los resultados con ambos criterios coinciden sólo presentaremos los correspondientes al primer criterio.

El experimento se ha realizado en dos etapas. En la primera se ha efectuado un pequeño experimento exploratorio, para determinar que variables experimentales son importantes y cuales no parecen tener efecto significativo sobre los resultados. En la segunda se ha realizado un estudio detallado sobre las variables experimentales importantes determinadas en la primera. Como resultado de la primera etapa se diseña el experimento que se indica en la tabla 1. Se consideran 12 modelos posibles de generación de un vector de seis series temporales. Estos modelos resultan de combinar: (1) dos ecuaciones de generación de las series a partir de los factores (una con dos factores y otra con tres); (2) tres ecuaciones de generación de factores (factores I(0), factores I(1) y factores I(2)); (3) dos estructuras posibles para las varianzas de las innovaciones de los factores (que hacen que éstos tengan igual o distinta varianza). Cuando los factores comunes se generan con igual varianza las matrices de covarianzas de las innovaciones de los factores son la identidad, y cuando los factores tienen distinta varianza las matrices son las que figuran en la tabla 1. En todos los casos la matriz de covarianzas de las innovaciones de las series es la identidad. Para cada uno de estos 12 modelos se han generado 500 vectores de series temporales de 100 observaciones. En cada replicación se generan vectores de seis variables y 200 observaciones, de las cuales se desprecian las 100 primeras, sobre la base del modelo factorial siguiente:

$$X_t = Af_t + u_t,$$

$$(1\text{-}B)^d f_t = a_t, \ d = 0,1,2$$

siendo A la matriz de carga de dimensión 6*r, f_t el vector de factores comunes, y u_t, a_t las innovaciones de las variables y de los factores comunes respectivamente, Σ_u, Σ_a las matrices de covarianzas de estas innovaciones. Posteriormente, estimamos la matriz factorial, asociada al vector generado, empleando cada uno de los tres métodos propuestos y obtenemos el error cuadrático de la estimación de acuerdo con (5). La estimación de la matriz de carga se realiza suponiendo que el número de factores comunes es conocido e igual al verdadero valor utilizado en la generación y tomando por tanto dos o tres vectores propios de cada matriz asociados a los dos o tres valores propios mayores. La medida de la "bondad", para cada método de estimación, es el promedio de los errores cuadráticos obtenidos en las 500 replicaciones. En las tablas 2 y 3 presentamos los valores medios con dos y tres factores y las desviaciones típicas de las distribuciones de 500 replicaciones, que es una medida adicional de la fiabilidad del método: cuanto menor sea esta desviación típica más predecibles son los resultados.

Resultados de la simulación

Las tablas 2-A y 3-A indican que el mejor procedimiento para estimar la matriz de carga, de acuerdo con el criterio del mínimo error cuadrático medio, es en general el del modelo factorial de Peña y Box (1987) y Peña y Poncela (1998). El resultado es consistente, tanto para dos como para tres factores. La diferencia es particularmente notable cuando los factores comunes son estacionarios, I(0). Esto es esperable con el método de Johansen, que está diseñado para factores I(1), pero sorprende el mal funcionamiento del análisis canónico en estos casos. El error de estimación aumenta mucho en los tres métodos cuando los factores son I(2), y entonces el mejor procedimiento, con una pequeña diferencia, es el de Johansen. Las tablas 2-A y 3-A indican una ventaja del método factorial respecto a su variabiliadad con relación a los otros dos métodos considerados. El método de Box y Tiao presenta de manera sistemática un error mayor que los otros dos.

Análisis de Robustez de los resultados

Para investigar como los resultados anteriores se modifican si, como suele ocurrir con series reales, las series contienen valores atípicos se han introducido atípicos de la forma siguiente: (1) El número de atípicos en cada replicación se obtiene generando un valor al azar de una normal con media y desviación típica igual a cinco. El número de atípicos resulta de redondear el valor anterior al entero más próximo. (2) Cada atípico se inserta en una sola de las series observadas, que se elige tomando un valor al azar entre 1 y 6. (3) La posición del atípico dentro de la serie se selecciona al azar mediante una distribución uniforme entre (1,100). (4) El atípico se introduce en la posición seleccionada con un valor de cinco desviaciones típicas de la serie de mayor variabilidad.

Las tablas 2-B y 3-B presentan los resultados cuando las series están contaminadas con atípicos. Ahora las ventajas a favor del método factorial disminuyen y el método de Johansen parece ser más robusto a la presencia de contaminación, salvo en el caso en que los factores son estacionarios. La variabilidad es menor en general con el método de Johansen.

5. COMPARACIÓN MEDIANTE SIMULACIONES DE LOS MÉTODOS DE ESTIMACIÓN DEL COEFICIENTE DE COINTEGRACIÓN

Al aumentar la dimensión del vector de series, y consecuentemente el número de relaciones de cointegración, es esperable un mejor comportamiento de los modelos factoriales, que determinan el espacio factorial de baja dimensión directamente, frente a los modelos basados en cointegración, que requieren la estimación de un espacio de dimensión alta. Para comparar estos métodos en un diseño que pueda considerarse especialmente favorable para los métodos basados en cointegración hemos seleccionado el modelo generador de datos de Bewley y otros (1994), descrito en la tabla 4. Estos autores consideran un sistema bivariante en el que existe una relación de cointegración (por tanto un factor común) y efectúan una comparación entre el estimador del coeficiente de cointegración obtenido por el método de Johansen (1988) y el obtenido por el método de Box y Tiao (1977).

Como el estimador máximo verosímil del coeficiente de cointegración es asintóticamente insesgado y con distribución normal (véase Phillips, 1991 y Johansen, 1988, 1992), el criterio de comparación utiliza estas propiedades. En primer lugar se analiza el sesgo en la estimación del coeficiente de cointegración, y en segundo, la desviación de la normalidad en la estimación, medida por la asimetría y el apuntamiento de la distribución del estimador.

Bewley y otros (1994) realizan un experimento con seis factores determinados por los valores de los parámetros asociados al modelo generador de datos (tabla 4). Estos factores son los siguientes:

– Exogeneidad y endogeneidad de la tendencia común del modelo generador de datos;

– Existencia o no de deriva en el modelo generador de datos;

– Ruidos del modelo generador de datos con o sin correlación;

– Varianza del ruido asociado al modelo generador de datos;

– Tamaño muestral en las series generadas;

– Valor del coeficiente de corrección de error asociado al modelo generador de datos.

Hemos repetido el diseño de estos autores incluyendo ahora el método de estimación de factores comunes. El vector de cointegración se obtiene como el vector ortogonal a la estimación del vector de cargas factoriales. En cada simulación se efectúan 2000 replicaciones y para cada replicación se estima el coeficiente de cointegración. Una vez realizadas las 2.000 replicaciones se calculan:

– El sesgo medio de estimación del coeficiente de cointegración. El sesgo se calcula con la mediana de las 2000 estimaciones obtenidas. Sea $Me(\beta^*)$ esta mediana y β^* el estimador del coeficiente de cointegración β. Entonces $Sesgo(\beta^*) = Me(\beta^*) - \beta$.

– Los intervalos centrados en la mediana que contienen el 50%, 95%, y 99% de las observaciones en la estimación del coeficiente de cointegración. Llamaremos I_{50}, I_{95} e I_{99} a estos intervalos.

5.1. Análisis del sesgo de estimación del coeficiente de cointegración

En modelos sin deriva y con tendencia común débilmente exógena, la tabla 5-A, muestra que en general la mejor estimación es la dada por Johansen (1988). En el caso en que la muestra sea pequeña y el coeficiente de corrección de error alto el estimador de Box y Tiao (1977) tiene menor sesgo que el de Johansen (1988). El sesgo cometido con el método de Peña y Box (1987) es bastante mayor que el de los otros dos métodos analizados, en especial cuando se trata de muestras pequeñas y en el caso en que el coeficiente de corrección de error sea alto. En Peña y Box (1987), el sesgo crece a medida que incrementa el valor de θ, es decir, a medida que aumenta la correlación entre los ruidos. Si se consideran **modelos sin deriva y con tendencia común endógena**, tabla 5-B, en general el sesgo aumenta respecto a la situación de tendencia débilmente exógena, siendo el patrón de comportamiento del sesgo semejante en ambas. El procedimiento de Peña y Box (1987) en tendencia común endógena, cuando no existe correlación en el ruido, tiene un sesgo similar al de los otros dos métodos e, incluso, en algún caso es menor.

Para **modelos con deriva**, tabla 6, mejora sensiblemente el sesgo de los estimadores analizados respecto a los sin deriva. Es interesante resaltar que en modelos con deriva y tendencia común endógena, el sesgo del estimador de Peña y Box (1987) es en la mayoría de los casos del mismo orden de magnitud que el de Box y Tiao (1977) y el de Johansen (1988).

5.2. Apuntamiento de la distribución del estimador del coeficiente de cointegración

Siguiendo el método de Bewley y otros (1994) para evaluar el apuntamiento se utilizan los cocientes I_{95}/I_{50} e I_{99}/I_{50}, cuyos valores en la distribución Normal son 2,90 y 3,82, respectivamente. Los resultados obtenidos en nuestras simulaciones para los métodos de Box y Tiao (1977) y Johansen (1988) son muy similares a los de Bewley y otros (1994), (pueden verse en Manzanedo,1997), por lo que presentamos solamente en este trabajo los obtenidos con el método de Peña y Box (1987) . Siguiendo a Bewley y otros (1994) en la exposición de los resultados separaremos el caso de ruido no correlacionado y el de ruido correlacionado.

Ruido no correlacionado

La distribución del estimador del coeficiente de cointegración para **modelos sin deriva**, tabla 7-A, presenta un apuntamiento superior al de la distribución Normal. El apuntamiento crece en función directa de la velocidad de ajuste y con los métodos de Box y Tiao (1977) y Johansen (1988) alcanza valores muy elevados cuando el coeficiente de corrección de error es alto (Φ = 0,9) y el tamaño muestral es pequeño o mediano (N = 75 ó N = 150). El tamaño muestral no afecta significativamente al apuntamiento de la distribución, salvo cuando el coeficiente de corrección de error es alto. En **modelos sin deriva**, la varianza asociada a los residuos no tiene influencia importante sobre el apuntamiento en los métodos de Box y Tiao (1977) y Johansen (1988). Por el contrario, en el método de Peña y Box (1987) el apuntamiento se reduce de forma significativa cuando la varianza de los residuos es mayor de 0,5, el coeficiente de corrección de error alto y el tamaño muestral pequeño o mediano. En concreto, para muestras pequeñas si la varianza de los residuos varía de 0,5 a 2 el apuntamiento pasa de 51,36 a 5,48 .

En **modelos con deriva**, tabla 7-B, el apuntamiento asociado a la distribución del coeficiente de cointegración se reduce, si la varianza residual no es alta, con respecto al caso del modelo sin deriva. La reducción es muy importante cuando la muestra es de tamaño pequeño o mediano y la velocidad de ajuste igual a 0,9. **En modelos con deriva**, para la estimación con el método de Box y Tiao (1977) y el de Johansen (1988), concluimos:

– La influencia de la velocidad de ajuste sobre el apuntamiento no es muy importante, excepto para muestras pequeñas, coeficiente de corrección de error alto y varianza residual alta.

– El tamaño muestral no afecta significativamente al apuntamiento de la distribución, excepto cuando el coeficiente de corrección de error y la varianza residual son altos.

Con el método de Peña y Box (1987) en **modelos con deriva**, el tamaño muestral, el coeficiente de corrección de error y la varianza tienen muy poca influencia sobre el apuntamiento. La novedad respecto de los otros dos métodos es que no se presenta el problema de apuntamiento alto en muestras pequeñas, varianza alta y coeficiente de corrección de error alto. En el resto de los casos, el apuntamiento asociado al estimador de Peña y Box (1987) es muy parecido al de los otros dos métodos.

Ruido correlacionado

En modelos sin deriva y tendencia común débilmente exógena, tabla 8-A, el patrón de comportamiento para el apuntamiento es muy semejante al del caso de ruido no correlacionado. En todas las situaciones el apuntamiento, asociado a la distribución del estimador del coeficiente de cointegración, está por encima del de la distribución Normal. En los métodos de Box y Tiao (1977) y Johansen (1988), el apuntamiento asociado no depende del parámetro que regula el grado de correlación entre los ruidos. Para el método de Peña y Box (1987), la dependencia es apreciable cuando la velocidad de ajuste es alta (Φ=0,9). **En modelos sin deriva y tendencia común endógena**, tabla 8-B, se incrementa el apuntamiento respecto al caso débilmente exógena si se consideran muestras de tamaño pequeño y, también, en el caso en que la velocidad de ajuste sea alta, en el resto, el apuntamiento es muy similar. En cuanto al rango interfractil de la distribución, en el caso de tendencia común endógena es aproximadamente el doble que en el homólogo de débilmente exógena. Para Peña y Box (1987) el apuntamiento es muy sensible, al parámetro que regula la correlación, si el tamaño muestral es pequeño y velocidad de ajuste es baja.

En modelos con deriva en los tres métodos, tabla 9, el apuntamiento y el rango interfractil disminuyen respecto al caso de sin deriva, especialmente para el caso de velocidad de ajuste alta. Además, el apuntamiento es muy similar en todos los casos. El rango interfractil es aproximadamente el doble en tendencia común endógena que en débilmente exógena.

6. CONCLUSIONES

En este trabajo examinamos tres métodos para la estimación de factores comunes y cointegrados en vectores de series temporales. Se ha comprobado mediante simulación que si tenemos un vector de datos de dimensión apreciable y un número de factores no muy alto con relación al número de datos el procedimiento de estimación de factores comunes de Peña y Box (1997) generalizado por Peña y Poncela (1998) parece funcionar mejor que los métodos de Box y Tiao y Johansen, aunque parece menos robusto que el de Johansen cuando la serie tiene datos atípicos. Por el contrario, cuando el número de series es pequeño y se desea estimar una relación de cointegración el método de Johansen tiene un sesgo menor que los dos métodos anteriores, si bien cuando el proceso generador de datos tiene deriva, los tres tienen un comportamiento similar.

REFERENCIAS

BARNEJEE, A. DOLADO, J., GALBRAITH, J. W. Y HENDRY, D. (1993). *Cointegration, Error Correction and the Econometric Analysis of Time Series*. Oxford University Press.

BEWLEY, R., ORDEN, D., YANG, M. Y FISHER, L. (1994). "Comparison of Box-Tiao and Johansen Canonical Estimators of Cointegrating Vectors in VEC(1) Models". *Journal of Econometrics*, 64, 3-27.

BOX, G.E.P. Y TIAO, G. C. (1977). "A Canonical Analysis of Múltiple Time Series". *Biometrica*, 64, 355-365.

CHAN, N. H. Y TSAY, R. S. (1991). "On the Use of Canonical Correlation Analysis in Testing Common Trends". Technical Report, Department of Statistics, Carnegie Mellon University.

CHAN, N. H. Y WEI, C. Z. (1988). "Limiting Distribution of Least Squares Estimates of Unstable Autorregresive Processes". *The Annals of Statistics*, 16, 367-401.

ENGLE, R.F. Y GRANGER, C.W.J. (1987). "Co-integration and Error Correction : Representation, Estimation and Testing". *Econometrica*, 55, 251-276.

ESCRIBANO, A. Y PEÑA, D. (1994). "Cointegration and Common Factors". *Journal of Time Series Analysis*, 15, 6, 577-586.

GONZALO, J. (1994). "Five Alternative Methods of Estimating Long-Run Equilibrium Relationships". *Journal of Econometrics*, 60, 203-233.

GONZALO, J. Y GRANGER, C.W.J. (1995). "Estimation of Common Long-Memory Components in Cointegrated Systems". *Journal of Business and Economics Statistics*, 13, 27-35.

JOHANSEN, S. (1988). "Statistical Analysis of Cointegrating Vectors". *Journal of Economic Dynamics y Control*, Vol. 12, 231-254.

JOHANSEN, S. (1992). "A Representation of Vector Autorregresive Processes Integrated of Order 2". *Econometric Theory*, 8, 188-202.

JOHANSEN, S. (1995). "A Statistical Analysis of Cointegration for I(2) Variables". *Econometric Theory*, 11, 25-59.

HARGREAVES, C: (1994). “A Review of Methods of Estimating Cointegrating Relationships”, In Hargreaves, C. (ed.) *Nonstationary Time Series Analysis and Cointegration, pp.87-131* Oxford University Press.

MANZANEDO, L. (1997)."Factores Comunes en Series Temporales Múltiples: Análisis del Crecimiento Europeo 1950/1990". *Tesis Doctoral*. Universidad Autónoma de Madrid.

PEÑA, D. Y BOX, G. E. P. (1987). "Identifying a Simplifying Structure in Time Series". *Journal of the American Statistical Association*, 82, 836-843.

PEÑA, D. Y PONCELA, P. (1998). "The Eigenstructure of Dynamic Factor Models ". *Universidad Carlos III de Madrid Working Paper*.

PHILLIPS, P. (1991). "Optimal Inference in Cointegration Systems". *Econometrica*, 59, 283-306.

QUENOULLI, M. (1957). *The Analysis of Multiple Time Series*. London:Griffin.

REINSEL, G. (1993). *Elements of Multivariate Time Series Analysis*. Springer-Verlag.

REINSEL, G. Y AHN, S. (1992). "VAR Models With Unit Roots and Reduced Rank Structure: Estimation, Likelihood Ratio Test and Forecasting". *Journal of Time Series Analysis*, 13, 4, 353-375.

TIAO, G. C. Y TSAY, R. S. (1989). "Model Specifications in Multivariate Time Series (with discussion)". *Journal of the Royal Statistical Society, Series B*, 51, 157-213.

STOCK, J. H. Y WATSON, M. W. (1988). "Testing for Common Trends". *Journal of the American Statistical Assotiation*, 83 1097-1107.

STOCK, J. H. Y WATSON, M. W. (1993). "A simple Estimator of Cointegrating Vector in Higher OrderIntegrated System". *Econometrica*, 61, 783-820.

APÉNDICE

Tabla 1

DISEÑO DEL MODELO FACTORIAL

Variables experimentales

A:Integración factores comunes	I(0)	I(1)	I(2)
B:Tamaño muestral	100		
C:Número factores comunes	2	3	
D:Varianza factores comunes	igual	diferente	
E:Número replicaciones	500		

Matriz factorial

Dos factores comunes

MGD(1) I

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

MGD(1) II

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0{,}5 & 0{,}5 \\ 0{,}25 & 0{,}75 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Tres factores comunes

MGD(1) I

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

MGD(1) II

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0{,}5 & 0{,}5 & 0 \\ 0 & 0{,}5 & 0{,}5 \\ 0{,}5 & 0 & 0{,}5 \end{bmatrix}$$

Varianza de las innovaciones de los factores comunes

Dos factores comunes

$$\begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix}$$

Tres factores comunes

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 8 \end{bmatrix}$$

(1) Modelo generador de datos

TABLA 2-A

ERROR CUADRÁTICO MEDIO DE LA ESTIMACIÓN Y SU DESVIACIÓN TÍPICA EN EL EXPERIMENTO DE SIMULACIÓN PARA LOS TRES MÉTODOS DE ESTIMACIÓN DE FACTORES COMUNES CON EL MODELO DE GENERACIÓN DE DATOS I. SE HA REMARCADO EL MENOR VALOR. MEDIAS DE 500 REPLICACIONES PARA SEIS VARIABLES DE TAMAÑO 100

	Dos factores comunes			*Tres factores comunes*		
	Box-Tiao	*Peña-Box*	*Johansen*	*Box-Tiao*	*Peña-Box*	*Johansen*
ERROR CUADRÁTICO MEDIO						
Diferente varianza en el ruido de los factores I (0)	0,96	**0,19**	0,76	1,18	**0,24**	0,92
Diferente varianza en el ruido de los factores I (1)	0,04	**0,03**	0,04	0,09	**0,04**	0,07
Diferente varianza en el ruido de los factores I (2)	2,01	1,85	**1,68**	2,43	2,37	**2,22**
Igual varianza en el ruido de los factores I (0)	0,62	**0,11**	0,51	0,55	**0,09**	0,48
Igual varianza en el ruido de los factores I (1)	0,03	**0,02**	**0,02**	0,03	**0,02**	0,03
Igual varianza en el ruido de los factores I (2)	2,03	1,86	**1,70**	2,48	2,33	**2,26**
DESVIACIÓN TÍPICA						
Diferente varianza en el ruido de los factores I (0)	0,54	**0,08**	0,51	0,50	**0,08**	0,50
Diferente varianza en el ruido de los factores I (1)	0,03	**0,01**	0,03	0,03	**0,02**	0,05
Diferente varianza en el ruido de los factores I (2)	0,65	0,63	**0,59**	**0,65**	**0,65**	0,68
Igual varianza en el ruido de los factores I (0)	0,62	**0,11**	0,57	0,66	**0,14**	0,63
Igual varianza en el ruido de los factores I (1)	0,04	**0,02**	0,04	0,14	**0,04**	0,11
Igual varianza en el ruido de los factores I (2)	0,64	0,62	**0,57**	0,67	**0,66**	0,69

TABLA 2-B

ERROR CUADRÁTICO MEDIO DE LA ESTIMACIÓN Y SU DESVIACIÓN TÍPICA EN EL EXPERIMENTO DE SIMULACIÓN PARA LOS TRES MÉTODOS DE ESTIMACIÓN DE FACTORES COMUNES CON EL MODELO DE GENERACIÓN DE DATOS I CUANDO LAS SERIES ESTÁN CONTAMINADAS POR ATÍPICOS. SE HA REMARCADO EL MENOR VALOR . MEDIAS DE 500 REPLICACIONES PARA SEIS VARIABLES DE TAMAÑO 100

	Dos factores comunes			*Tres factores comunes*		
	Box-Tiao	*Peña-Box*	*Johansen*	*Box-Tiao*	*Peña-Box*	*Johansen*
ERROR CUADRÁTICO MEDIO						
Diferente varianza en el ruido de los factores I (0)	1,14	0,98	**0,84**	1,20	1,03	**0,94**
Diferente varianza en el ruido de los factores I (1)	0,62	0,87	**0,37**	0,82	1,27	**0,59**
Diferente varianza en el ruido de los factores I (2)	2,10	1,89	**1,76**	2,55	**2,24**	2,29
Igual varianza en el ruido de los factores I (0)	1,33	**0,79**	1,02	1,55	**0,90**	1,19
Igual varianza en el ruido de los factores I (1)	0,40	0,79	**0,23**	0,74	1,20	**0,42**
Igual varianza en el ruido de los factores I (2)	2,10	1,85	**1,77**	2,52	**2,23**	2,25
DESVIACIÓN TÍPICA						
Diferente varianza en el ruido de los factores I (0)	0,67	0,85	**0,60**	0,68	0,92	**0,64**
Diferente varianza en el ruido de los factores I (1)	0,70	0,89	**0,51**	0,74	1,04	**0,61**
Diferente varianza en el ruido de los factores I (2)	0,67	0,78	**0,62**	**0,68**	0,87	0,69
Igual varianza en el ruido de los factores I (0)	0,64	0,73	**0,63**	0,71	0,79	**0,68**
Igual varianza en el ruido de los factores I (1)	0,54	0,85	**0,34**	0,73	0,98	**0,49**
Igual varianza en el ruido de los factores I (2)	0,65	0,81	**0,56**	**0,62**	0,85	0,70

TABLA 3-A

ERROR CUADRÁTICO MEDIO DE LA ESTIMACIÓN Y SU DESVIACIÓN TÍPICA EN EL EXPERIMENTO DE SIMULACIÓN PARA LOS TRES MÉTODOS DE ESTIMACIÓN DE FACTORES COMUNES CON EL MODELO DE GENERACIÓN DE DATOS II. SE HA REMARCADO EL MENOR VALOR. MEDIAS DE 500 REPLICACIONES PARA SEIS VARIABLES DE TAMAÑO 100

	Dos factores comunes			*Tres factores comunes*		
	Box-Tiao	*Peña-Box*	*Johansen*	*Box-Tiao*	*Peña-Box*	*Johansen*
ERROR CUADRÁTICO MEDIO						
Diferente varianza en el ruido de los factores I (0)	1,87	**0,12**	0,56	2,48	**0,09**	0,57
Diferente varianza en el ruido de los factores I (1)	1,78	**0,02**	**0,02**	2,49	**0,02**	0,03
Diferente varianza en el ruido de los factores I (2)	2,39	1,84	**1,66**	3,01	2,44	**2,21**
Igual varianza en el ruido de los factores I (0)	2,08	**0,21**	0,72	2,65	**0,24**	0,92
Igual varianza en el ruido de los factores I (1)	1,78	**0,03**	**0,03**	2,61	**0,04**	0,05
Igual varianza en el ruido de los factores I (2)	2,46	1,87	**1,69**	2,93	2,40	**2,24**
DESVIACIÓN TÍPICA						
Diferente varianza en el ruido de los factores I (0)	0,52	**0,10**	0,55	0,64	**0,07**	0,59
Diferente varianza en el ruido de los factores I (1)	0,46	**0,02**	**0,02**	0,70	**0,02**	**0,02**
Diferente varianza en el ruido de los factores I (2)	0,70	0,69	**0,66**	**0,64**	0,66	0,67
Igual varianza en el ruido de los factores I (0)	0,62	**0,14**	0,60	0,73	**0,22**	0,63
Igual varianza en el ruido de los factores I (1)	0,52	**0,02**	0,03	0,79	**0,03**	0,04
Igual varianza en el ruido de los factores I (2)	0,66	0,55	**0,59**	0,77	**0,59**	0,64

TABLA 3-B

ERROR CUADRÁTICO MEDIO DE LA ESTIMACIÓN Y SU DESVIACIÓN TÍPICA EN EL EXPERIMENTO DE SIMULACIÓN PARA LOS TRES MÉTODOS DE ESTIMACIÓN DE FACTORES COMUNES CON EL MODELO DE GENERACIÓN DE DATOS II CUANDO LAS SERIES ESTÁN CONTAMINADAS POR ATÍPICOS. SE HA REMARCADO EL MENOR VALOR . MEDIAS DE 500 REPLICACIONES PARA SEIS VARIABLES DE TAMAÑO 100

	Dos factores comunes			*Tres factores comunes*		
	Box-Tiao	*Peña-Box*	*Johansen*	*Box-Tiao*	*Peña-Box*	*Johansen*
ERROR CUADRÁTICO MEDIO						
Diferente varianza en el ruido de los factores I (0)	1,96	0,86	**0,84**	2,66	**0,88**	1,02
Diferente varianza en el ruido de los factores I (1)	1,87	0,77	**0,20**	2,48	1,20	**0,50**
Diferente varianza en el ruido de los factores I (2)	2,40	1,76	**1,67**	2,87	2,39	**2,25**
Igual varianza en el ruido de los factores I (0)	2,18	**0,53**	0,87	2,79	**0,66**	1,11
Igual varianza en el ruido de los factores I (1)	1,80	0,53	**0,13**	2,55	0,99	**0,26**
Igual varianza en el ruido de los factores I (2)	2,48	1,84	**1,77**	2,88	2,45	**2,33**
DESVIACIÓN TÍPICA						
Diferente varianza en el ruido de los factores I (0)	**0,55**	**0,55**	0,62	0,73	**0,62**	0,69
Diferente varianza en el ruido de los factores I (1)	0,49	0,67	**0,30**	0,80	0,74	**0,54**
Diferente varianza en el ruido de los factores I (2)	**0,62**	0,66	0,68	0,70	**0,64**	0,68
Igual varianza en el ruido de los factores I (0)	**0,54**	0,47	0,62	0,73	**0,45**	0,67
Igual varianza en el ruido de los factores I (1)	0,53	0,64	**0,18**	0,76	0,66	**0,36**
Igual varianza en el ruido de los factores I (2)	0,61	**0,58**	**0,58**	**0,64**	0,66	**0,64**

TABLA 4

MODELO GENERADOR DE DATOS DE BEWLEY Y OTROS (1994)

$$y_t - \beta x_t = z_t \; ; \qquad z_t = \Phi z_{t-1} + e_{zt} \; ;$$

$$x_t + \alpha y_t = w_t \; ; \qquad w_t = \delta + w_{t-1} + e_{wt} \; ;$$

$$\text{var} = \begin{bmatrix} e_{zt} \\ \\ e_{wt} \end{bmatrix} = \begin{bmatrix} 1 & \theta\sigma \\ \\ \theta\sigma & \sigma^2 \end{bmatrix}$$

los valores utilizados para los parámetros son

$\delta = (0 ; 0{,}5);$ $\alpha = (0 ; 1);$ $\beta = 1;$

$\theta = (-0{,}5 ; -0{,}25 ; 0 ; 0{,}25 ; 0{,}5) ;$

$\sigma = (0{,}5 ; 1 ; 2);$ $\Phi = (0{,}3 ; 0{,}6 ; 0{,}9);$

$N = (75 ; 150 ; 300);$

Los valores de los parámetros asociados al proceso generador de datos se fijan para obtener un diseño que:

a.- Contempla modelos con tendencia común estocástica estrictamente exógena ($\alpha = 0$, $\theta = 0$), débilmente exógena ($\alpha = 0$, $\theta \neq 0$), y endógena ($\alpha = 1$, θ cualquiera).

b.- Se considera modelos con deriva ($\delta \neq 0$) y sin deriva ($\delta = 0$).

c.- Se tienen tres varianzas generadoras de la tendencia común (σ).

d.- Existen tres velocidades de ajuste al equilibrio de la relación de cointegración (Φ).

e.- Analiza tres tamaños muestrales diferentes, muestras pequeñas (N=75), muestras medianas (N=150) y muestras grandes (N=300).

TABLA 5

SESGO DE LOS ESTIMADORES EN LA MEDIANA PARA MODELOS SIN DERIVA (δ =0) Y VARIANZA DE LA PERTURBACIÓN LA UNIDAD (σ =1,0)

Parte A: Modelos con tendencia común débilmente exógena (α=0)

	Estimador Johansen			*Estimador Box y Tiao*			*Estimador Peña-Box*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
θ =-0,50									
75	-0,0017	0,0022	-0,0714	0,0221	0,0221	-0,0455	0,0563	0,0710	0,1192
N = 150	-0,0003	0,0012	-0,0004	0,0122	0,0116	0,0191	0,0286	0,0341	0,0922
300	0,0003	0,0017	0,0015	0,0067	0,0059	0,0056	0,0149	0,0168	0,0448
θ =-0,25									
75	-0,0013	-0,0018	-0,0198	0,0113	0,0073	-0,0088	0,0467	0,0623	0,1535
N = 150	0,0002	-0,0005	-0,0096	0,0058	0,0038	-0,0046	0,0238	0,0238	0,0765
300	0,0009	-0,0003	0,0047	0,0032	0,0016	0,0071	0,0129	0,0152	0,0495
θ =0,00									
75	0,0005	0,0031	-0,0089	0,0021	0,0023	-0,0068	0,0633	0,0824	0,2203
N = 150	0,0004	0,0009	0,0013	-0,0003	0,0018	-0,0024	0,0305	0,0431	0,1279
300	-0,0003	0,0011	-0,0036	-0,0002	0,0015	-0,0039	0,0147	0,0226	0,0675
θ =0,25									
75	0,0006	0,0021	0,0243	-0,0099	-0,0075	0,0146	0,0839	0,1259	0,3242
N = 150	0,0007	-0,0006	0,0118	-0,0054	-0,0054	0,0053	0,0433	0,0649	0,2063
300	0,0004	0,0006	-0,0021	-0,0023	-0,0024	-0,0062	0,0234	0,0349	0,1051
θ =0,50									
75	0,0007	0,0038	0,0401	-0,0255	-0,0169	0,0115	0,1442	0,2091	0,5595
N = 150	0,0016	0,0016	0,0148	-0,0109	-0,0086	0,0024	0,0783	0,1093	0,3371
300	-0,0002	-0,0006	0,0013	-0,0059	-0,0049	-0,0019	0,0364	0,0554	0,1811

TABLA 5 (Continuación)
SESGO DE LOS ESTIMADORES EN LA MEDIANA PARA MODELOS SIN DERIVA (δ =0) Y VARIANZA DE LA PERTURBACIÓN LA UNIDAD (σ =1,0)

Parte B : Modelos con tendencia común endógena (α=1)

	Estimador Johansen			Estimador Box y Tiao			Estimador Peña-Box		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
θ =-0,50									
75	-0,0026	-0,0005	-0,3034	0,0481	0,0341	-0,2670	-0,1064	-0,1763	-0,7187
N = 150	-0,0005	0,0012	-0,0405	0,0248	0,0204	-0,0280	-0,0463	-0,0809	-0,3673
300	0,0002	0,0006	-0,0105	0,0127	0,0108	-0,0022	-0,0236	-0,0408	-0,1799
θ =-0,25									
75	0,0018	-0,0005	-0,2336	0,0252	0,0086	-0,2107	-0,0439	-0,0809	-0,3995
N = 150	0,0003	-0,0011	-0,0422	0,0111	0,0069	-0,0313	-0,0209	-0,0364	-0,1702
300	0,0002	-0,0001	-0,0072	0,0056	0,0039	0,0014	-0,0097	-0,0187	-0,0724
θ =0,00									
75	0,0045	-0,0014	-0,1675	0,0051	-0,0011	-0,1402	0,0051	-0,0021	-0,1483
N = 150	-0,0003	0,0002	-0,0291	0,0002	0,0015	-0,0175	0,0004	0,0011	-0,0222
300	-0,0004	0,0021	0,0039	-0,0003	0,0021	0,0032	-0,0003	0,0019	0,0073
θ =0,25									
75	0,0014	0,0018	-0,1461	-0,0213	-0,0106	-0,1109	0,0459	0,0796	0,0830
N = 150	0,0014	0,0030	-0,0195	-0,0095	-0,0064	-0,0176	0,0229	0,0409	0,1366
300	0,0005	-0,0004	0,0031	-0,0049	-0,0044	0,0021	0,0107	0,0174	0,0817
θ =0,50									
75	-0,0007	0,0051	-0,1591	-0,0456	-0,0358	-0,1432	0,1082	0,1918	0,0519
N = 150	0,0019	0,0020	-0,0075	-0,0219	-0,0182	-0,0183	0,0499	0,0895	0,2901
300	-0,0005	0,0007	-0,0001	-0,0125	-0,0087	-0,0081	0,0231	0,0426	0,1798

TABLA 6

SESGO DE LOS ESTIMADORES EN LA MEDIANA PARA MODELOS CON DERIVA (δ=0,5) Y VARIANZA DE LA PERTURBACIÓN LA UNIDAD (σ=1,0)

Parte A: Modelos con tendencia común débilmente exógena (α=0)

	Estimador Johansen			*Estimador Box y Tiao*			*Estimador Peña-Box*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
θ=-0,50									
75	0,0008	-0,0004	0,0016	0,0022	0,0004	0,0000	0,0039	0,0027	0,0053
N = 150	0,0000	-0,0001	0,0004	0,0005	0,0001	0,0013	0,0007	0,0007	0,0023
300	0,0001	0,0000	-0,0001	0,0001	0,0001	-0,0002	0,0002	0,0002	0,0003
θ =-0,25									
75	-0,0001	0,0003	0,0020	0,0007	0,0010	0,0016	0,0033	0,0045	0,0097
N = 150	-0,0001	-0,0001	0,0004	0,0001	0,0001	0,0003	0,0008	0,0011	0,0024
300	0,0001	0,0002	-0,0002	0,0001	0,0001	-0,0002	0,0002	0,0004	0,0006
θ =0,00									
75	0,0004	-0,0002	-0,0012	0,0000	0,0003	-0,0001	0,0043	0,0062	0,0141
N = 150	0,0002	-0,0001	0,0001	0,0002	0,0000	0,0014	0,0013	0,0015	0,0022
300	-0,0001	-0,0001	-0,0001	-0,0001	-0,0002	-0,0005	0,0002	0,0003	0,0007
θ =0,25									
75	0,0002	0,0002	0,0033	-0,0004	-0,0002	0,0032	0,0063	0,0094	0,0240
N = 150	-0,0002	-0,0003	-0,0002	-0,0003	-0,0003	-0,0005	0,0013	0,0022	0,0058
300	0,0001	-0,0001	-0,0004	0,0000	-0,0001	-0,0006	0,0004	0,0006	0,0013
θ =0,50									
75	-0,0004	0,0014	-0,0012	-0,0016	-0,0005	0,0004	0,0072	0,0122	0,0268
N = 150	0,0001	-0,0002	-0,0009	-0,0002	-0,0004	-0,0020	0,0021	0,0031	0,0078
300	0,0000	-0,0001	0,0002	-0,0001	-0,0003	0,0001	0,0006	0,0005	0,0029

TABLA 6 (Continuación)

SESGO DE LOS ESTIMADORES EN LA MEDIANA PARA MODELOS CON DERIVA (δ=0,5) Y VARIANZA DE LA PERTURBACIÓN LA UNIDAD (σ=1,0)

Parte B :modelos con tendencia común endógena (α =1)

	Estimador Johansen			*Estimador Box y Tiao*			*Estimador Peña-Box*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
θ =-0,50									
75	-0,0007	-0,0018	0,0092	0,0035	0,0019	0,0107	-0,0042	-0,0086	-0,0167
N = 150	-0,0003	-0,0001	-0,0045	0,0001	0,0002	-0,0017	-0,0016	-0,0024	-0,0076
300	-0,0002	-0,0001	0,0011	0,0001	-0,0001	0,0001	-0,0005	-0,0007	-0,0025
θ =-0,25									
75	-0,0003	0,0012	-0,0090	0,0010	0,0023	0,0008	-0,0032	-0,0034	-0,0103
N = 150	-0,0001	0,0011	-0,0004	0,0003	0,0014	0,0011	-0,0009	-0,0002	-0,0018
300	0,0001	0,0001	-0,0006	0,0003	0,0001	-0,0008	0,0000	-0,0003	-0,0022
θ =0,00									
75	0,0001	-0,0006	-0,0025	0,0004	0,0004	0,0036	0,0003	0,0002	0,0037
N = 150	-0,0002	-0,0001	0,0000	-0,0002	-0,0005	-0,0020	-0,0002	-0,0004	-0,0028
300	-0,0001	-0,0003	0,0000	-0,0001	0,0001	0,0007	-0,0001	-0,0001	0,0009
θ =0,25									
75	0,0003	0,0017	-0,0086	-0,0018	-0,0018	-0,0127	0,0025	0,0034	-0,0006
N = 150	0,0005	0,0010	0,0013	0,0000	0,0006	0,0016	0,0012	0,0019	0,0052
300	0,0001	0,0001	0,0008	-0,0001	0,0000	-0,0003	0,0002	0,0004	0,0004
θ =0,50									
75	-0,0003	0,0012	-0,0020	-0,0025	-0,0045	-0,0072	0,0050	0,0062	0,0207
N = 150	0,0002	-0,0003	-0,0030	-0,0008	0,0010	-0,0029	0,0012	0,0015	0,0034
300	-0,0002	0,0000	-0,0011	-0,0002	-0,0002	-0,0016	-0,0003	-0,0005	0,0006

TABLA 7

RANGO INTERFRACTIL Y APUNTAMIENTO ASOCIADOS AL ESTIMADOR DE PEÑA Y BOX PARA MODELOS CON TENDENCIA COMÚN ESTRICTAMENTE EXÓGENA

(α =0; θ =0)

Parte A : Modelos sin deriva (δ =0)

	I99/I50			*I95/I50*			*I50*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
σ =0,5									
75	6,22	7,07	51,36	3,98	3,79	10,61	0,3481	0,5430	1,6591
N = 150	5,42	5,65	16,19	3,63	3,70	4,92	0,1677	0,2521	0,8847
300	5,81	5,02	5,39	3,66	3,62	3,56	0,0823	0,0366	0,0603
σ =1,0									
75	5,23	5,29	8,71	3,38	3,75	4,26	0,1022	0,1446	0,4471
N = 150	5,15	5,15	5,67	3,61	3,51	3,61	0,0485	0,0782	0,2533
300	5,94	5,32	5,04	3,45	3,46	3,49	0,7226	0,8550	0,1415
σ =2,0									
75	5,02	5,39	5,48	3,63	3,56	3,39	0,0366	0,0603	0,1721
N = 150	5,44	5,19	4,87	3,39	3,49	3,24	0,0197	0,0311	0,1113
300	5,44	5,35	4,65	3,59	3,76	3,35	0,0099	0,0154	0,0586

Parte B : Modelos con deriva (δ =0,5)

	I99/I50			*I95/I50*			*I50*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
σ =0,5									
75	4,14	4,15	3,99	3,06	3,03	2,95	0,0198	0,0336	0,1124
N = 150	3,91	3,73	4,01	2,91	2,88	2,87	0,0072	0,0126	0,0458
300	3,99	3,82	3,83	2,96	2,86	2,93	0,0025	0,0046	0,0169
σ =1,0									
75	5,37	5,07	4,87	3,39	3,17	3,16	0,0192	0,0345	0,1131
N = 150	4,27	4,51	4,34	3,06	3,11	3,15	0,0072	0,0118	0,0445
300	4,01	4,12	3,76	2,95	3,02	2,71	0,0026	0,0044	0,0181
σ =2,0									
75	7,81	6,88	5,64	4,31	4,09	3,56	0,0185	0,0315	0,1066
N = 150	6,53	5,41	5,87	3,62	3,38	3,53	0,0068	0,0124	0,0442
300	4,59	4,51	4,42	3,16	3,19	3,11	0,0026	0,0047	0,0177

TABLA 8

RANGO INTERFRACTIL Y APUNTAMIENTO ASOCIADOS AL ESTIMADOR DE PEÑA Y BOX PARA MODELOS SIN DERIVA ($\delta = 0$) Y VARIANZA DE LA PERTURBACIÓN LA UNIDAD ($\sigma = 1$)

Parte A : Modelos con tendencia común débilmente exógena ($\alpha = 0$)

	I99/I50			*I95/I50*			*I50*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
θ =-0,50									
75	6,43	7,48	31,59	3,81	4,06	5,77	0,1112	0,1715	0,6481
N = 150	5,82	5,93	10,62	3,68	3,81	4,34	0,0546	0,0836	0,3365
300	5,42	4,95	5,30	3,53	3,65	3,64	0,0278	0,0418	0,1569
θ =-0,25									
75	5,42	6,05	10,91	3,46	3,71	4,69	0,0966	0,1534	0,5044
N = 150	5,38	5,48	6,13	3,66	3,88	4,04	0,0464	0,0681	0,2512
300	5,08	4,82	5,45	3,48	3,49	3,54	0,0237	0,0376	0,1407
θ =0,00									
75	5,23	5,29	8,71	3,38	3,75	4,26	0,1022	0,1446	0,4471
N = 150	5,15	5,15	5,67	3,61	3,51	3,61	0,0485	0,0782	0,2533
300	5,94	5,32	5,04	3,45	3,46	3,49	0,7226	0,8550	0,1415
θ =0,25									
75	4,96	4,47	7,45	3,35	3,12	3,81	0,1261	0,1891	0,4772
N = 150	5,01	4,84	4,62	3,47	3,56	3,34	0,0638	0,0966	0,3128
300	4,99	4,84	4,57	3,36	3,13	3,36	0,0311	0,0509	0,1684
θ =0,50									
75	4,58	4,48	6,11	3,44	3,21	3,67	0,1772	0,2584	0,6636
N = 150	5,26	4,52	4,78	3,62	3,22	3,14	0,0934	0,1420	0,4148
300	4,62	4,95	4,59	3,35	3,26	3,28	0,0456	0,0739	0,2385

TABLA 8 (Continuación)
RANGO INTERFRACTIL Y APUNTAMIENTO ASOCIADOS AL ESTIMADOR DE PEÑA Y BOX PARA MODELOS SIN DERIVA ($\delta=0$) Y VARIANZA DE LA PERTURBACIÓN LA UNIDAD ($\sigma=1$)

Parte B : Modelos con tendencia común endógena (α=1)

	I99/I50			*I95/I50*			*I50*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
θ =-0,50									
75	7,08	8,96	38,97	4,12	3,97	7,77	0,2332	0,4003	0,8825
N = 150	6,10	5,81	24,77	3,76	3,71	4,71	0,1054	0,1799	0,6498
300	5,94	5,46	6,46	3,69	3,59	3,69	0,0521	0,0936	0,3612
θ =-0,25									
75	7,56	9,98	42,39	4,05	4,31	8,59	0,1718	0,2901	0,9058
N = 150	5,81	5,99	34,18	3,72	3,79	6,23	0,0821	0,1412	0,5355
300	5,64	5,32	7,67	3,72	3,59	3,94	0,0393	0,6960	0,2726
θ =0,00									
75	9,57	13,71	70,28	4,69	4,91	11,74	0,1541	0,2707	1,0606
N = 150	6,65	6,55	41,66	3,82	3,88	7,62	0,0734	0,1293	0,5526
300	5,41	5,54	10,45	3,69	3,84	4,63	0,0371	0,0629	0,2579
θ =0,25									
75	10,86	25,73	89,34	4,68	6,11	15,49	0,1904	0,3586	1,5610
N = 150	6,61	7,74	62,20	3,94	4,31	12,11	0,0848	0,1536	0,7274
300	5,79	5,91	15,42	3,74	3,81	5,19	0,0396	0,0712	0,3207
θ =0,50									
75	40,28	92,35	86,01	7,31	16,09	14,24	0,2816	0,5673	3,5771
N = 150	8,37	16,62	123,84	4,49	5,33	21,61	0,1197	0,2249	1,3151
300	6,11	6,33	61,40	3,79	3,93	8,17	0,0547	0,0981	0,5334

TABLA 9

RANGO INTERFRACTIL Y APUNTAMIENTO ASOCIADOS AL ESTIMADOR DE PEÑA Y BOX PARA MODELOS CON DERIVA (δ =0,5) Y VARIANZA DE LA PERTURBACIÓN LA UNIDAD (σ =1)

Parte A : Modelos con tendencia común debilmente exógena (α =0)

	I99/I50			*I95/I50*			*I50*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
θ =-0,50									
75	4,67	4,52	4,69	3,15	3,24	3,28	0,0226	0,0377	0,1254
N = 150	4,16	4,39	4,07	2,95	3,14	3,07	0,0085	0,0133	0,0489
300	3,80	3,74	3,97	2,99	2,88	2,92	0,0029	0,0052	0,0195
θ =-0,25									
75	4,88	4,72	4,31	3,25	3,19	3,30	0,0203	0,0357	0,1193
N = 150	4,28	4,16	4,29	3,10	3,03	3,17	0,0072	0,0129	0,0465
300	4,05	3,97	4,12	2,87	3,00	3,00	0,0027	0,0046	0,0178
θ =0,00									
75	4,58	4,72	4,70	3,09	3,12	3,25	0,0200	0,0338	0,1106
N = 150	4,16	4,23	4,34	3,03	3,05	3,06	0,0072	0,0119	0,0454
300	3,94	4,05	4,01	2,96	3,00	3,14	0,0025	0,0043	0,0170
θ =0,25									
75	4,69	4,46	4,76	3,31	3,31	3,23	0,0196	0,0350	0,1130
N = 150	4,19	3,91	4,08	3,20	2,94	3,08	0,0073	0,0134	0,0468
300	4,41	4,10	4,04	3,23	2,89	2,99	0,0025	0,0047	0,0179
θ =0,50									
75	4,24	4,64	4,41	3,23	3,20	3,09	0,0225	0,0384	0,1194
N = 150	3,86	4,32	3,91	2,95	3,07	3,01	0,0082	0,0136	0,0495
300	4,13	3,77	3,93	3,13	2,96	2,93	0,0027	0,0049	0,0188

TABLA 9 (Continuación)
RANGO INTERFRACTIL Y APUNTAMIENTO ASOCIADOS AL ESTIMADOR DE PEÑA Y BOX PARA MODELOS CON DERIVA ($\delta = 0{,}5$) Y VARIANZA DE LA PERTURBACIÓN LA UNIDAD ($\sigma = 1$)

Parte B : Modelos con tendencia común endógena ($\alpha = 1$)

	I99/I50			*I95/I50*			*I50*		
Φ	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*	*0,3*	*0,6*	*0,9*
θ =-0,50									
75	4,87	4,71	4,75	3,29	3,49	3,25	0,0426	0,0722	0,2524
N = 150	4,29	3,95	4,43	3,20	2,99	3,10	0,0155	0,0278	0,0982
300	4,04	4,39	3,82	2,94	3,09	2,93	0,0055	0,0098	0,0385
θ =-0,25									
75	5,10	4,42	5,39	3,31	3,14	3,38	0,0391	0,0701	0,2380
N = 150	4,45	4,33	4,15	3,22	3,05	2,89	0,0141	0,0258	0,0810
300	4,07	3,95	4,28	3,08	3,03	2,96	0,0051	0,0091	0,0355
θ =0,00									
75	4,78	4,79	5,33	3,33	3,31	3,43	0,0400	0,0701	0,2313
N = 150	4,40	4,17	4,34	3,01	2,98	3,06	0,0142	0,0250	0,0911
300	3,96	3,88	3,90	2,99	2,96	2,96	0,0051	0,0090	0,0341
θ =0,25									
75	4,96	4,74	6,02	3,23	3,32	3,48	0,0419	0,0698	0,2334
N = 150	4,48	4,48	4,13	2,99	3,10	3,12	0,0147	0,0252	0,0963
300	3,89	4,04	3,89	2,85	2,96	3,02	0,0054	0,0094	0,0365
θ =0,50									
75	4,33	4,60	5,05	3,19	3,10	3,50	0,0434	0,0779	0,2581
N = 150	4,17	4,55	4,14	2,91	3,24	3,04	0,0160	0,0276	0,1037
300	3,83	3,89	3,69	3,06	2,88	2,95	0,0056	0,0100	0,0380

A COMPARISON BETWEEN METHODS OF ESTIMATION OF COMMON FACTORS AND COINTEGRATION RELATIONSHIPS IN VECTORS OF TIME SERIES

SUMMARY

We compare in a Monte Carlo study the procedures proposed by Box and Tiao (1977), Peña and Box (1987) and Johansen (1988) for the estimation of common and cointegrated factors in time series. When the data are generated by a factor model the three methods work well when estimating common factors I(1)and are less efficient for the case of I(2). Overall, the best procedure seems to be the one by Peña and Box. In the estimation of a cointegration coefficient the best overall method is the one due to Johansen although when the data generating process has a drift the three methods present similar properties.

Keywords: Eigenvectors, linear transformations, nonstationary time series.

AMS Classification: 90A20, 62M20, 62P20.